# Quantitative Methods – II

A.Y. 2021-22

## Practice 13

Lorenzo Cavallo

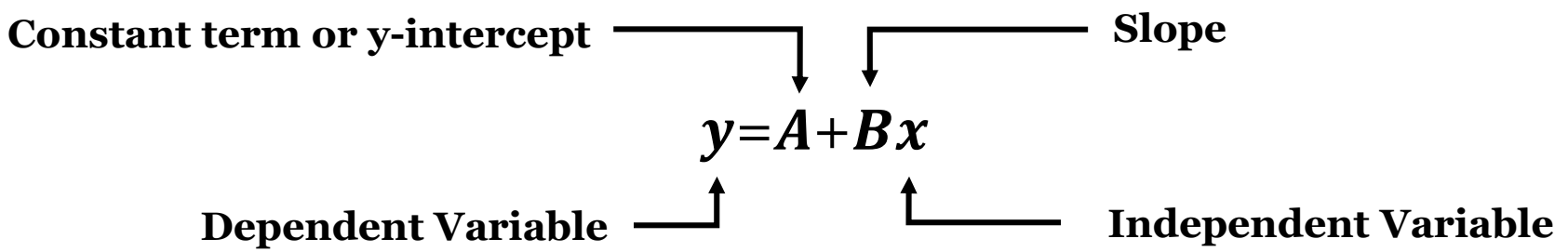
For any clarification/meeting: cavallo@istat.it

# THEME #1

# Simple Linear Regression

## Simple Linear Regression

**Regression** is a statistical procedure that determines the equation for the straight line that best fits a specific set of data.
Any straight line can be represented by an equation of the form




where ***A*** and ***B*** are constants.

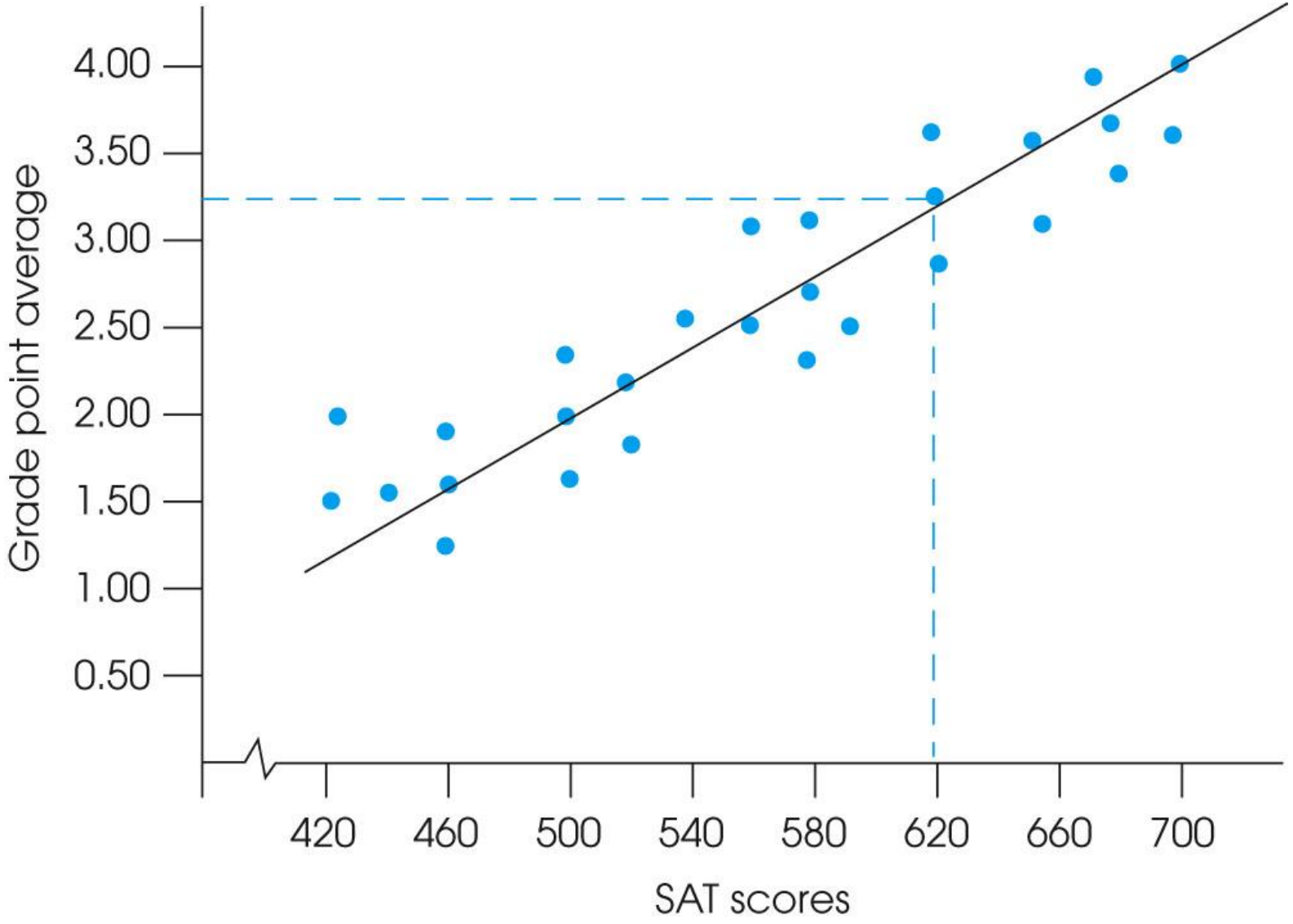
The value of ***A*** is called the Y-intercept and determines the point where the line crosses the Y-axis.

The value of ***B*** is called the slope constant and determines the direction and degree to which the line is tilted.

Each set of values of ***A*** and ***B*** gives a different straight line.

The regression line is also called ***the regression of y on x***

Hypothetical data showing the relationship between SAT scores and GPA with a regression line drawn through the data points. The **regression line** defines a precise, one-to-one relationship between each ***x*** value (SAT score) and its corresponding ***y*** value (GPA).



In case of samples, we have to estimate the regression model from sample data by the equation:

$$\hat{\boldsymbol{y}} = \boldsymbol{a} + \boldsymbol{bx}$$

Where:

- $\hat{\boldsymbol{y}}$ is the estimated or predicted value of $\boldsymbol{y}$ for a given value of $\boldsymbol{x}$
- $\boldsymbol{a}$ is the estimator of the *Y-intercept* $\boldsymbol{A}$
- $\boldsymbol{b}$ is the estimated value of the *slope* $\boldsymbol{B}$.

The differences between the observed values $\boldsymbol{y_i}$ and the estimated values $\hat{\boldsymbol{y}}_{\boldsymbol{i}}$ are called residuals $\boldsymbol{e_i}$

$$e_i = (y_i - \hat{y}_i)$$

The **error sum of square** is

$$SSE = \sum (y_i - \hat{y}_i)^2 = \sum {e_i}^2$$

To calculate the estimated values $\boldsymbol{a}$ of the Y-intercept and $\boldsymbol{b}$ of the slope we have to find the values that give the minimum *SSE*

This estimated values are called the least squares estimates of $\boldsymbol{A}$ and $\boldsymbol{B}$:

$$\boldsymbol{b} = \frac{\boldsymbol{SS_{xy}}}{\boldsymbol{SS_{xx}}}$$

where *SS* stands for "sum of squares"

$$SS_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n} \qquad \textit{Covariance (x,y)}$$

$$SS_{xx} = \sum x^2 - \frac{(\sum x)^2}{n} \qquad \textit{Variance of x}$$

$$\boldsymbol{a} = \overline{\boldsymbol{y}} - \boldsymbol{b}\overline{\boldsymbol{x}}$$

where $\overline{\boldsymbol{x}}$ and $\overline{\boldsymbol{y}}$ are the means of $\boldsymbol{x}$ and $\boldsymbol{y}$

## Exercise 1

The following information is obtained from a sample data set.

$$n = 10, \qquad \sum x = 100, \qquad \sum y = 220, \qquad \sum xy = 3680, \qquad \sum x^2 = 1140$$

Find the estimated regression line.

Solution

Data are:

$$n = 10, \qquad \sum x = 100, \qquad \sum y = 220, \qquad \sum xy = 3680, \qquad \sum x^2 = 1140$$

The estimated regression line is:

$$\hat{y} = a + b\,x$$

We have to calculate the estimated values $a$ and $b$ for the regression line where:

$$a = \bar{y} - b\bar{x}$$

$$b = \frac{SS_{xy}}{SS_{xx}}$$

Where:

$\bar{x}$ and $\bar{y}$ are the means of x and y

$SS_{xy}$ is the Covariance (x,y): $SS_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n}$

$SS_{xx}$ the Variance of x: $SS_{xx} = \sum x^2 - \frac{(\sum x)^2}{n}$

## Exercise 1

The following information is obtained from a sample data set.

$$n = 10, \qquad \sum x = 100, \qquad \sum y = 220, \qquad \sum xy = 3680, \qquad \sum x^2 = 1140$$

Find the estimated regression line.

Solution

The means of x and y are:

$$\bar{x} = \frac{\sum x}{n} = \frac{100}{10} = 10 \quad \text{and} \quad \bar{y} = \frac{\sum y}{n} = \frac{220}{10} = 22$$

The Covariance (x;y) is:

$$SS_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n} = 3680 - \left(\frac{100 \cdot 220}{10}\right) = 3680 - 2200 = 1480$$

The Variance of x is:

$$SS_{xx} = \sum x^2 - \frac{(\sum x)^2}{n} = 1140 - \frac{100^2}{10} = 140$$

So the slope «$b$» will be:

$$b = \frac{SS_{xy}}{SS_{xx}} = \frac{1480}{140} = 10.5714$$

And the Y-intecept «$a$»:

$$a = \bar{y} - b\bar{x} = 22 - 10.5714 \cdot 10 = -83.7143$$

The regression line will be:

$$\hat{y} = -83.71 + 10{,}57\, x$$

## Exercise 2

The following information is obtained from a sample data set.

$$n = 12, \qquad \sum x = 66, \qquad \sum y = 588, \qquad \sum xy = 2244, \qquad \sum x^2 = 396$$

Find the estimated regression line.

Solution

Data:

$$n = 12, \qquad \sum x = 66, \qquad \sum y = 588, \qquad \sum xy = 2244, \qquad \sum x^2 = 396$$

To calculate $a$ and $b$ we have to calculate the means, the covariance and the variance of x

$$\bar{x} = \frac{\sum x}{n} = \frac{66}{12} = 5.5 \quad \text{and} \quad \bar{y} = \frac{\sum y}{n} = \frac{588}{12} = 49$$

The Covariance (x;y) is:

$$SS_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n} = 2244 - \left(\frac{66 \cdot 588}{12}\right) = -990$$

The Variance of x is:

$$SS_{xx} = \sum x^2 - \frac{(\sum x)^2}{n} = 396 - \frac{66^2}{12} = 33$$

So the slope «$b$» and the will be Y-intecept «$a$»:

$$b = \frac{SS_{xy}}{SS_{xx}} = \frac{-990}{33} = -30$$

$$a = \bar{y} - b\bar{x} = 49 + 30 \cdot 5.5 = 214$$

The regression line will be:

$$\hat{y} = 214 - 30\,x$$

## Exercise 3

The following information is obtained from a sample data set.

$$n = 15, \qquad \sum x = 25, \qquad \sum y = -150, \qquad \sum xy = -2500, \qquad \sum x^2 = 550$$

Find the estimated regression line

Solution

Data:

$$n = 15, \qquad \sum x = 25, \qquad \sum y = -150, \qquad \sum xy = -2500, \qquad \sum x^2 = 550$$

To calculate $a$ and $b$ we have to calculate the means, the covariance and the variance of x

$$\bar{x} = \frac{\sum x}{n} = \frac{25}{15} = 1.67 \quad \text{and} \quad \bar{y} = \frac{\sum y}{n} = \frac{-150}{15} = -10$$

The Covariance (x;y) is:

$$SS_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n} = -2500 - \left(\frac{25 \cdot -150}{15}\right) = -2250$$

The Variance of x is:

$$SS_{xx} = \sum x^2 - \frac{(\sum x)^2}{n} = 550 - \frac{25^2}{15} = 508.33$$

So the slope «$b$» and the will be Y-intecept «$a$»:

$$b = \frac{SS_{xy}}{SS_{xx}} = \frac{-2250}{508.33} = -4.43$$

$$a = \bar{y} - b\bar{x} = -10 + 4.43 \cdot 1.67 = -2.62$$

The regression line will be:

$$\hat{y} = -2.62 - 4{,}43\, x$$

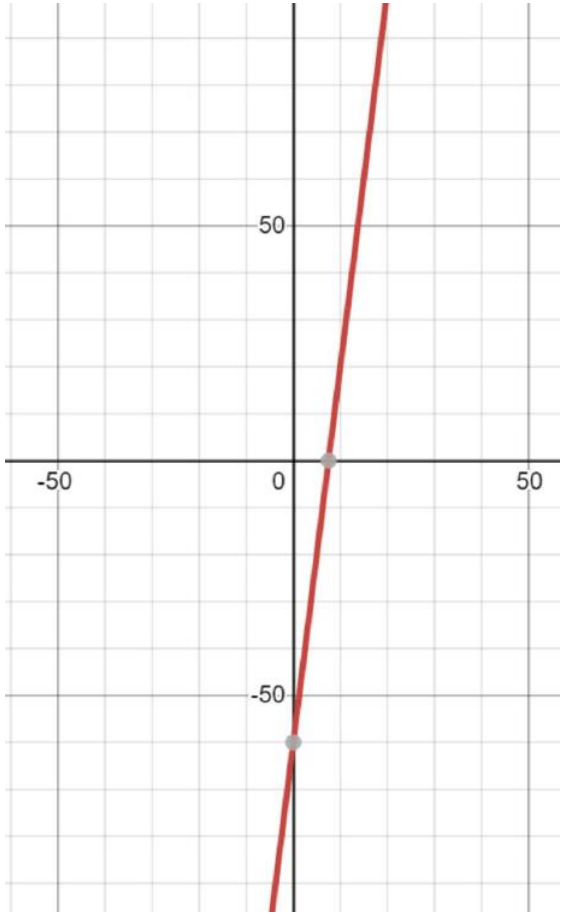
## Exercise 4

Plot the following straight lines. Give the values of the y-intercept and slope for each of these lines and interpret them. Indicate whether each of the lines gives a positive or a negative relationship between x and y.

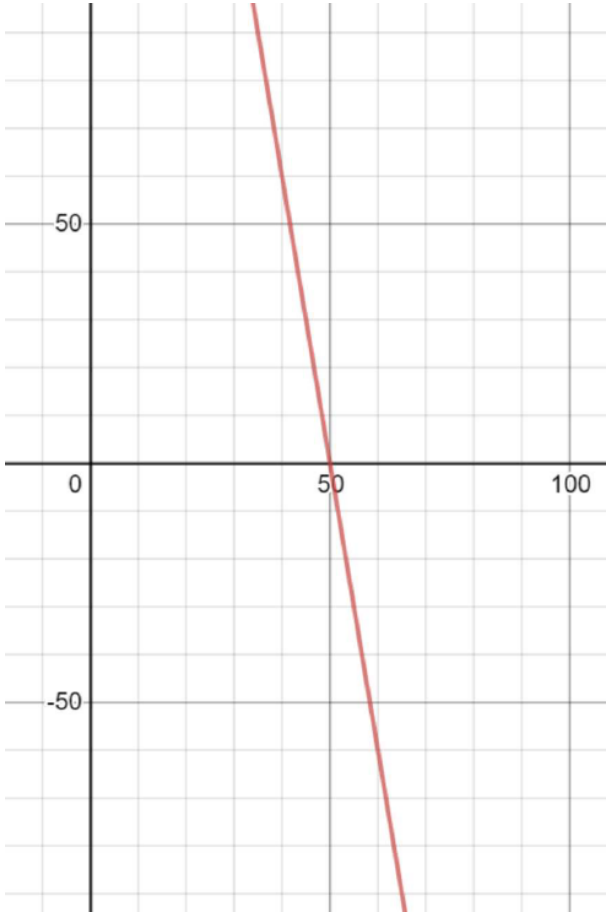
a. $y = -60 + 8x$
b. $y = 300 - 6x$
c. $y=0.5x + 15$

Solution

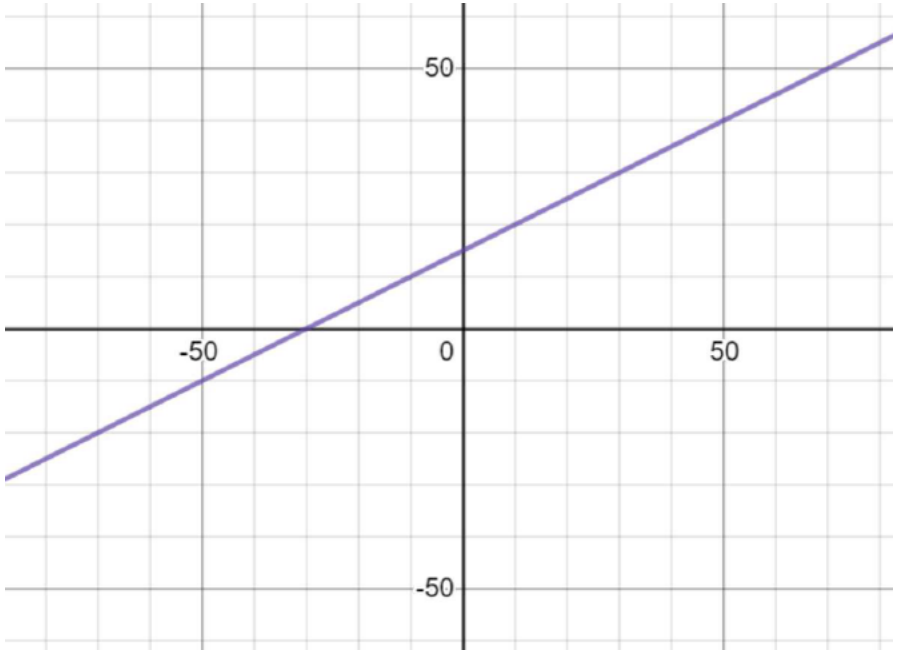
$y = -60 + 8x$




$y = 300 - 6x$




$y=0.5x + 15$



## Exercise 5

We observed 8 values for the variables X and Y:

| $x_i$ | $y_i$ |
|---|---|
| -2 | 2 |
| -5 | -3 |
| 4 | 10 |
| 5 | 8 |
| 8 | 20 |
| 10 | 60 |
| -7 | -18 |
| 12 | 24 |

Using x as independent variable and y as dependent, find the least square regression line.

Solution

| x | y | x · y | $x^2$ |
|---|---|---|---|
| -2 | 2 | -4 | 4 |
| -5 | -3 | 15 | 25 |
| 4 | 10 | 40 | 16 |
| 5 | 8 | 40 | 25 |
| 8 | 20 | 160 | 64 |
| 10 | 60 | 600 | 100 |
| -7 | -18 | 126 | 49 |
| 12 | 24 | 288 | 144 |
| 25 | 103 | 1265 | 427 |

$$\bar{x} = \frac{\sum x}{n} = \frac{25}{8} = 3.125 \quad \text{and} \quad \bar{y} = \frac{\sum y}{n} = \frac{103}{8} = 12.875$$

$$SS_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n} = 1265 - \left(\frac{25 \cdot 103}{8}\right) = 943.125$$

$$SS_{xx} = \sum x^2 - \frac{(\sum x)^2}{n} = 427 - \frac{25^2}{8} = 348.875$$

$$b = \frac{SS_{xy}}{SS_{xx}} = \frac{943.125}{348.875} = 2.7$$

$$a = \bar{y} - b\bar{x} = 12.875 - 2.7 \cdot 3.125 = 4.43$$

$$\hat{y} = 4.43 + 2{,}7\, x$$

## Exercise 6

We observed 6 values for the variables X and Y:

| $x_i$ | $y_i$ |
|---|---|
| 12 | 1 |
| 4 | 13 |
| 5 | 16 |
| 8 | 8 |
| 6 | 4 |
| 10 | 3 |

Using x as independent variable and y as dependent, find the least square regression line.

Solution

| x | y | x · y | $x^2$ |
|---|---|---|---|
| 12 | 1 | 12 | 144 |
| 4 | 13 | 52 | 16 |
| 5 | 16 | 80 | 25 |
| 8 | 8 | 64 | 64 |
| 6 | 4 | 24 | 36 |
| 10 | 3 | 30 | 100 |
| 45 | 45 | 262 | 385 |

$$\bar{x} = \frac{\sum x}{n} = \frac{45}{6} = 7.5 \quad \text{and} \quad \bar{y} = \frac{\sum y}{n} = \frac{45}{6} = 7.5$$

$$SS_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n} = 262 - \left(\frac{45 \cdot 45}{6}\right) = -75.5$$

$$SS_{xx} = \sum x^2 - \frac{(\sum x)^2}{n} = 385 - \frac{45^2}{6} = 47.5$$

$$b = \frac{SS_{xy}}{SS_{xx}} = \frac{-75.5}{47.5} = -1.59$$

$$a = \bar{y} - b\bar{x} = 7.5 + 1.59 \cdot 7.5 = 19.42$$

$$\hat{y} = 19.42 - 1.59\,x$$

**The standard deviation of error**

$$s_e = \sqrt{\frac{SS_{yy} - bSS_{xy}}{n-2}}$$

where $SS_{yy} = \sum y^2 - \frac{(\sum y)^2}{n}$ *Variance of y*

The **Total Sum of Squares** (***SST***) is:

$$SST = \sum y^2 - \frac{(\sum y)^2}{n} = SS_{yy}$$

In general

$$SST = SSR + SSE$$

Where:
***SSR*** is the **regression Sum of Squares**
***SSE*** is the **error Sum of Squares**

The **Regression Sum of Squares** (***SSR***) is:

$$SSR = SST - SSE = \sum y^2 - \frac{(\sum y)^2}{n} - \sum (y - \hat{y})^2 = \sum (\hat{y} - \bar{y})^2$$

The **Total Sum of Squares** (***SST***)

$$SST = SS_{yy} = \sum (y - \bar{y})^2 = SSR + SSE = \sum (\hat{y} - \bar{y})^2 + \sum (y - \hat{y})^2$$

## Coefficient of Determination

The coefficient of determination, $R^2$ or $r^2$, represents the proportion of ***SST*** explained by the use of the regression model.

$$R^2 = \frac{b\ SS_{xy}}{SS_{yy}}$$

and $0 \leq R^2 \leq 1$

$$R^2 = \frac{SSR}{SST} = \frac{SST - SSE}{SST} = 1 - \frac{SSE}{SST}$$

## Exercise 7

In a regression analysis the error sum of squares is 150 and the regression sum of squares is 50. With this data calculate the coefficient of determination.

Solution
We can calculate the coefficient of determination as:

$$R^2 = \frac{SSR}{SST} = \frac{SST - SSE}{SST} = 1 - \frac{SSE}{SST}$$

We have the SSE=150 and the SSR=50, the total sum of squares SST is:

$$SST = SSR + SSE = 150 + 50 = 200$$

So,

$$R^2 = \frac{SSR}{SST} = \frac{SSR}{SSR + SSE} = \frac{50}{200} = 0.25 = 25\% \ (low\ correlation)$$

## Linear Correlation Coefficient

The simple linear correlation coefficient, ***r***, measures the strength of the linear relationship between two variables for a sample and is calculated as

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}}$$

where $-1 \leq r \leq 1$

The coefficient of determination ($R^2$) is the square of the correlation coefficient ($r$)

**The standard error of the regression ($s_e$)** represents the average distance that the observed values fall from the regression line.

$$s_e = \sqrt{\frac{SS_{yy} - bSS_{xy}}{n-2}}$$

where $SS_{yy} = \sum y^2 - \frac{(\sum y)^2}{n}$

*The reason n-2 is used rather than n-1 is that two parameters (the slope and the intercept) were estimated in order to estimate the sum of squares.*

## Exercise 8

Compute the standard deviation of errors, $s_e$, and the coefficient of determination, $r^2$, for the data of the exercise 5 and 6.

Solution for 5

| x | y | x · y | $x^2$ | $y^2$ |
|---|---|---|---|---|
| -2 | 2 | -4 | 4 | 4 |
| -5 | -3 | 15 | 25 | 9 |
| 4 | 10 | 40 | 16 | 100 |
| 5 | 8 | 40 | 25 | 64 |
| 8 | 20 | 160 | 64 | 400 |
| 10 | 60 | 600 | 100 | 3600 |
| -7 | -18 | 126 | 49 | 324 |
| 12 | 24 | 288 | 144 | 576 |
| 25 | 103 | 1265 | 427 | 5077 |

$$\bar{x} = \frac{\sum x}{n} = \frac{25}{8} = 3.125 \quad \text{and} \quad \bar{y} = \frac{\sum y}{n} = \frac{103}{8} = 12.875$$

$$SS_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n} = 1265 - \left(\frac{25 \cdot 103}{8}\right) = 943.125$$

$$SS_{xx} = \sum x^2 - \frac{(\sum x)^2}{n} = 427 - \frac{25^2}{8} = 348.875$$

$$b = \frac{SS_{xy}}{SS_{xx}} = \frac{943.125}{348.875} = 2.7$$

$$SS_{yy} = \sum y^2 - \frac{(\sum y)^2}{n} = 5077 - \frac{103^2}{8} = 3750.88$$

$$s_e = \sqrt{\frac{SS_{yy} - bSS_{xy}}{n-2}} = \sqrt{\frac{3750.88 - 2.7 \cdot 943.125}{6}} = 14.15$$

$$R^2 = \frac{b\ SS_{xy}}{SS_{yy}} = \frac{2.7 \cdot 943.125}{3750.88} = 0.68$$

## Exercise 8

Compute the standard deviation of errors, *se,* and the coefficient of determination, $r^2$, for the data of the exercise 5 and 6.

Solution for 6

| x | y | x · y | $x^2$ | $y^2$ |
|---|---|---|---|---|
| 12 | 1 | 12 | 144 | 1 |
| 4 | 13 | 52 | 16 | 169 |
| 5 | 16 | 80 | 25 | 256 |
| 8 | 8 | 64 | 64 | 64 |
| 6 | 4 | 24 | 36 | 16 |
| 10 | 3 | 30 | 100 | 9 |
| 45 | 45 | 262 | 385 | 515 |

$$\bar{x} = \frac{\sum x}{n} = \frac{45}{6} = 7.5 \quad \text{and} \quad \bar{y} = \frac{\sum y}{n} = \frac{45}{6} = 7.5$$

$$SS_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n} = 262 - \left(\frac{45 \cdot 45}{6}\right) = -75.5$$

$$SS_{xx} = \sum x^2 - \frac{(\sum x)^2}{n} = 385 - \frac{45^2}{6} = 47.5$$

$$b = \frac{SS_{xy}}{SS_{xx}} = \frac{-75.5}{47.5} = -1.59$$

$$SS_{yy} = \sum y^2 - \frac{(\sum y)^2}{n} = 515 - \frac{45^2}{6} = 177.5$$

$$s_e = \sqrt{\frac{SS_{yy} - bSS_{xy}}{n-2}} = \sqrt{\frac{177.5 - (-1.59) \cdot (-75.5)}{4}} = 3.79$$

$$R^2 = \frac{b\, SS_{xy}}{SS_{yy}} = \frac{-1.59 \cdot -75.5}{177.5} = 0.68$$

# THEME #1

## Inference about the slope

## Confidence Interval for *B*

The (1 – α)% confidence interval for B is given by

$$b \pm ts_b$$

where

$$s_b = \frac{s_e}{\sqrt{SS_{xx}}}$$

and the value of *t* is obtained from the t distribution table for α=α/2 area in the right tail of the t distribution and n – 2 degrees of freedom.

## Test Statistic for *b*

The value of the test statistic *t* for *b* is calculated as

$$t = \frac{b - B}{s_b}$$

The value of *B* is substituted from the null hypothesis.

## Exercise 9

a. Calculate the linear correlation coefficient for the Exercise 5, 6, 8 and 9.
b. Calculate for the Exercise 8, the Confidence interval of B (α=0.05) and test the hypothesis $H_0$: B=-1 vs $H_1$: B≠-1 (α=0.05)

Solution

a.

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}}$$

Exercise 5:

$$SS_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n} = 1265 - \left(\frac{25 \cdot 103}{8}\right) = 943.125$$

$$SS_{xx} = \sum x^2 - \frac{(\sum x)^2}{n} = 427 - \frac{25^2}{8} = 348.875$$

$$SS_{yy} = \sum y^2 - \frac{(\sum y)^2}{n} = 5077 - \frac{103^2}{8} = 3750.88$$

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}} = \frac{943.125}{\sqrt{348.875 \cdot 3750.88}} = 0.82$$

## Exercise 9

a. Calculate the linear correlation coefficient for the Exercise 5, 6, 8 and 9.
b. Calculate for the Exercise 8, the Confidence interval of B (α=0.05) and test the hypothesis $H_0$: B=-1 vs $H_1$: B≠-1 (α=0.05)

Solution

a.

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}}$$

Exercise 6:

$$SS_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n} = 262 - \left(\frac{45 \cdot 45}{6}\right) = -75.5$$

$$SS_{xx} = \sum x^2 - \frac{(\sum x)^2}{n} = 385 - \frac{45^2}{6} = 47.5$$

$$SS_{yy} = \sum y^2 - \frac{(\sum y)^2}{n} = 515 - \frac{45^2}{6} = 177.5$$

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}} = -0.82$$

## Exercise 9

a. Calculate the linear correlation coefficient for the Exercise 5, 6, 8 and 9.
b. Calculate for the Exercise 8, the Confidence interval of B ($\alpha$=0.05) and test the hypothesis $H_0$: B=-1 vs $H_1$: B≠-1 ($\alpha$=0.05)

Solution

a.

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}}$$

Exercise 8:

$$SS_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n} = -26$$

$$SS_{xx} = \sum x^2 - \frac{(\sum x)^2}{n} = 104$$

$$SS_{yy} = \sum y^2 - \frac{(\sum y)^2}{n} = 108.83$$

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}} = -0.24$$

## Exercise 9

a. Calculate the linear correlation coefficient for the Exercise 5, 6, 8 and 9.
b. Calculate for the Exercise 8, the Confidence interval of B (α=0.05) and test the hypothesis $H_0$: B=-1 vs $H_1$: B≠-1 (α=0.05)

Solution

a.

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}}$$

Exercise 9:

$$SS_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n} = 39.2$$

$$SS_{xx} = \sum x^2 - \frac{(\sum x)^2}{n} = 51.2$$

$$SS_{yy} = \sum y^2 - \frac{(\sum y)^2}{n} = 91.2$$

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}} = 0.57$$

## Exercise 9

a. Calculate the linear correlation coefficient for the Exercise 5, 6, 8 and 9.
b. Calculate for the Exercise 8, the Confidence interval of $B$ ($\alpha$=0.05) and test the hypothesis $H_0$: $B$=-1 vs $H_1$: $B \neq$-1 ($\alpha$=0.05)

Solution

b. Confidence interval for $b$:

$$b \pm ts_b$$

$$\text{where} \quad s_b = \frac{s_e}{\sqrt{SS_{xx}}} \text{ with } s_e = \sqrt{\frac{SS_{yy} - bSS_{xy}}{n-2}} \text{ and } t_{n\text{-}2,\alpha/2} \text{ (two tailed)}$$

Exercise 8:

$n$=6, $b$=-0.25 , $s_e = 5.06$ , $SS_{xx} = 104$,

Hence,

$$s_b = \frac{5.06}{\sqrt{104}} = 0.5 \text{ and } t_{\text{n-2},\alpha/2} = t_{4,0.025}=2.78$$

$$(1\text{-}\alpha)\%\text{IC}(B) = 95\%\text{IC}(B) = b \pm ts_b = [-0.25\text{-}2.78 \cdot 0.5;\ -0.25+2.78 \cdot 0.5] = [-1.64\ ;\ 1.14]$$

$$t = \frac{b - B}{s_b} = \frac{-0.25 + 1}{0.5} = 1.5 < 2.78$$

We do not reject the null hypothesis.

## Exercise 10

We have the following sample data for the variable X and Y

| $x_i$ | $y_i$ |
|---|---|
| 10 | 9.4 |
| 20 | 9.2 |
| 50 | 9.0 |
| 100 | 8.5 |
| 150 | 8.1 |
| 200 | 7.4 |

a. Calculate the correlation coefficient;
b. Estimate the parameter of the regression model

Solution a.

| $x_i$ | $y_i$ | $x_i^2$ | $y_i^2$ | $x_i y_i$ |
|---|---|---|---|---|
| 10 | 9.40 | 100 | 88.36 | 94 |
| 20 | 9.20 | 400 | 84.64 | 184 |
| 50 | 9.00 | 2500 | 81 | 450 |
| 100 | 8.50 | 10000 | 72.25 | 850 |
| 150 | 8.10 | 22500 | 65.61 | 1215 |
| 200 | 7.40 | 40000 | 54.76 | 1480 |
| **530** | **51.60** | **75500** | **446.62** | **4273** |

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}}$$

$$SS_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n} = 4273 - \frac{530 \cdot 51.6}{6} = -285$$

$$SS_{xx} = \sum x^2 - \frac{(\sum x)^2}{n} = 75500 - \frac{530^2}{6} = 28683.33$$

$$SS_{yy} = \sum y^2 - \frac{(\sum y)^2}{n} == 446.62 - \frac{51.6^2}{6} = 2.86$$

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}} = \frac{-285}{\sqrt{28683.33 \cdot 2.86}} = -0.995$$

## Exercise 10

Solution b.

| $x_i$ | $y_i$ | $x_i^2$ | $y_i^2$ | $x_i y_i$ |
|---|---|---|---|---|
| 10 | 9.40 | 100 | 88.36 | 94 |
| 20 | 9.20 | 400 | 84.64 | 184 |
| 50 | 9.00 | 2500 | 81 | 450 |
| 100 | 8.50 | 10000 | 72.25 | 850 |
| 150 | 8.10 | 22500 | 65.61 | 1215 |
| 200 | 7.40 | 40000 | 54.76 | 1480 |
| **530** | **51.60** | **75500** | **446.62** | **4273** |

$$\hat{y} = a + bx$$

$$a = \bar{y} - b\bar{x}$$

$$b = \frac{SS_{xy}}{SS_{xx}}$$

$$\bar{x} = \frac{\sum x}{n} = \frac{530}{6} = 88.33$$

$$\bar{y} = \frac{\sum y}{n} = \frac{51.6}{6} = 8.6$$

$$SS_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n} = -285 \; and \; SS_{xx} = \sum x^2 - \frac{(\sum x)^2}{n} = 28683.33$$

$$b = \frac{SS_{xy}}{SS_{xx}} = \frac{-285}{28683.33} = -0.009936$$

$$a = \bar{y} - b\bar{x} = 8.6 - (-0.009936) \cdot 88.3 = 9.4777$$

$$\hat{y} = 9.4777 - 0.009936x$$